

SELFALIGNED PROCESS FOR A FLASH MEMORY

5

FIELD OF THE INVENTION

The present invention relates generally to a semiconductor process, and more particularly to, a selfaligned process for a flash memory.

10

BACKGROUND OF THE INVENTION

In a complex integrated circuit (IC), the shrinkage of the devices thereof makes the design more difficult. Process such as selfalignment and other techniques are used for the desired designs.

15

FIG. 1 shows a cross-sectional view of a gate structure 10 in a typical flash memory, in which a tunnel oxide 14 is formed on a substrate 12 with a floating gate polysilicon layer 16 thereon, and an oxide-nitride-oxide (ONO) layer 18 is further formed on the polysilicon layer 16 with a control gate polysilicon layer 20 thereon. Moreover, a control gate

20

25

tungsten silicide layer 22 and a hard mask layer 24 are formed on the polysilicon layer 20, and a source 30 and a drain 32 are formed on the substrate 12. In the formation of the gate structure 10, deposition and etching processes are used to obtain a gate stack on the substrate 12. Then selfaligned process with the gate stack as a mask is used to form the source 30 and drain 32. After forming the spacers 26 and 28, a selfaligned process is further used to form source and drain contacts. Prior arts are proposed for such selfaligned processes, such as in U.S. pat. nos. 5,907,781 and 6,444,530 issued to Chen et al.

However, due to the tungsten silicide layer sandwiched in the gate structure, the critical dimension (CD) will be enlarged and the distances between the gate and the contact windows of the source and drain will be shortened by the thermal expansion of the tungsten silicide layer resulted from the thermal stress when the crystal structure of the tungsten silicide is transferred from tetragon cubic to hexagon cubic in the subsequent thermal process, such as oxidation and annealing. In addition, the breakdown voltage of the structure is lower for the shortage of spaces therebetween. Further, the re-growth of the tungsten silicide grain squeezes each other and causes the sidewall of the gate structure rough and uneven, and as a result, the local electric filed effect is

enhanced and induces unpredicted discharge at sharp corners to damage the gate structure and shorten the lifetime of the flash memory. It is therefore desired a selfaligned process to obtain gate structure having a flat sidewall for flash memories.

5

SUMMARY OF THE INVENTION

10 An object of the present invention is to provide a selfaligned process to reach a smoothed sidewall of the gate for a flash memory.

15 Another object of the present invention is to provide a selfaligned process to enhance the voltage endurance of the gate structure for a flash memory.

20 In a selfaligned process for a flash memory, according to the present invention, a first polysilicon layer, ONO layer, second polysilicon layer, tungsten silicide layer and hard mask layer are deposited on a tunnel oxide layer and etched to form a gate structure. Then the sidewall of the tungsten silicide layer is cleaned by a solution having a high etch selectivity to tungsten silicide after the formation of source and drain with the gate structure as a mask and before
25 annealing process. Spacer is further formed on the sidewall of

the gate structure, and selfaligned contact window process is subsequently preceded.

5

BRIEF DESCRIPTION OF DRAWINGS

These and other objects, features and advantages of the present invention will become apparent to those skilled in the art upon consideration of the following description of the preferred embodiments of the present invention taken in
10 conjunction with the accompanying drawings, in which:

FIG. 1 is the cross-sectional view of the gate structure in a typical flash memory;

15

FIG. 2 is the cross-sectional view after forming a gate stack and source/drain;

FIG. 3 is the cross sectional view after forming a spacer on the sidewall of the gate structure;

20

FIG.4 is an illustration of a conventional gate structure after deformation caused by thermal expansion of the tungsten silicide in the gate structure;

Fig. 5 is a microscope photograph of a conventional gate structure; and

25

Fig. 6 is a microscope photograph of the gate structure produced by the inventive process.

DETAILED DESCRIPTION OF THE INVENTION

An embodiment according to the present invention to illustrate a selfaligned process for a flash memory is shown in FIG. 2 to FIG. 3. FIG. 2 is the cross-sectional view after a gate stack 10 and a source 30/drain 32 are formed, in which a tunnel oxide layer 14 is formed on a substrate 12 with a polysilicon layer 16 thereon, an ONO layer 18 is formed on the polysilicon layer 16 with another polysilicon layer 20 thereon, and a tungsten silicide layer 22 is formed on the second polysilicon layer 20 with a hard mask layer 24 thereon. After forming the gate stack 10, the source 30 and drain 32 are formed on the substrate 12 with the gate structure 10 as the mask in the first selfaligned process. Then a solution with a high etch selectivity to tungsten silicide is used to clean the sidewall of the tungsten silicide layer 22 in the gate structure 10. Preferably, SC-1 is used for the solution in this clean process. SC-1 is an alkaline peroxide solution that composes of five parts of deionized water, one part of 30% hydrogen peroxide, and one part of 29% ammonia. After this step, the sidewall of the tungsten silicide layer 22 is etched for the control of the critical dimension of the tungsten silicide layer 22. When the high etching selectivity solution is used to clean the sidewall of the tungsten silicide layer 22, the

tungsten silicide layer 22 has a faster etching rate than other layers, and thus the sidewall of the tungsten silicide layer 22 is etched to form recess on its sidewall.

5 After the above clean process is finished, rapid thermal processing (RTP) is preceded in an atmosphere containing oxygen radical in a chamber so as to activate the gate, source and drain structures and form an oxide layer at the polysilicon 16 outskirts of the floating gate polysilicon 16 to
10 prevent current leakage. Due to the thermal treatment using rapid thermal treatment in an atmosphere with oxygen radical, surface reaction is the main reaction mechanism of the thermal oxidation in this atmosphere and thus the tungsten silicide layer 22 is kept smooth on its surface and not easy to
15 expand. When using the rapid thermal process in an atmosphere with oxygen free radical, hydrogen and oxygen are additionally pumped into the chamber at a low pressure from about 5 torrs to 50 torrs.

20 After the above annealing process, the crystal structure of the tungsten silicide layer 22 is transferred from tetragon cubic crystal to hexagon cubic crystal. SiN or SiO₂ is deposited and etched to form spacers 26 and 28 on the sidewalls of the gate structure 10, as shown in FIG. 3. Since
25 the tungsten silicide layer 22 is previously cleaned with a high

etch selectivity solution, gaps 34 and 36 are formed between the tungsten silicide layer 22 and spacers 26 and 29 from the recesses, and thereby increasing the distance between the tungsten silicide layer 22 and spacers 26 and 28. When the tungsten silicide layer 22 expands due to the thermal stress in the subsequent thermal process, no squeezing is happened to damage the gate structure 10. As a result, the surface of the tungsten silicide layer 22 and gate structure 10 is kept smooth, and the distances between the tungsten silicide layer 22 and contact windows will not be shortened. Poor performance such as increasing in local electric field and decreasing in breakage voltage won't happen.

FIG. 4 is an illustration of the deformation of a conventional gate structure due to the thermal expansion of the tungsten silicide for comparison with the resultant structure formed by the inventive process. In a conventional selfaligned process, the tungsten silicide layer in a gate structure will expand due to thermal stress in any subsequent thermal process. Since there is no excess space in the gate structure for the expanded volume of the tungsten silicide layer caused by thermal stress, the grains inside the tungsten silicide layer will push each other and increases the critical dimension of the tungsten silicide layer, and the breakage voltage between the tungsten silicide layer and contact

windows will be lowered. In contrast, the inventive selfaligned process generates a buffer gaps 34 and 36 between the tungsten silicide layer 22 and spacers 26 and 28. When the tungsten silicide layer 22 expands due to thermal stress, gaps 34 and 36 buffer the expansion of the tungsten silicide layer 22 and thereby do not affect the structure inside the tungsten silicide layer 22.

Fig. 5 is a microscope photograph of a conventional gate structure. The squeezing of the tungsten silicide layer due to the thermal expansion caused by thermal stress can be seen thereof, and the surface of the sidewall is very rough. Fig. 6 is a microscope photograph of the gate structure produced by the inventive process, in which the tungsten silicide layer is not squeezed due to the expansion caused by thermal stress, and the surface of the sidewall is very smooth. Comparing Figs. 5 and 6, the selfaligned process proposed in the present invention has obviously improved the disadvantages of the prior gate structure.

While the present invention has been described in conjunction with preferred embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended

to embrace all such alternatives, modifications and variations that fall within the spirit and scope thereof as set forth in the appended claims.